Special_Issue
ICLR_2022_v1
pdf

Visual Concepts

- Solution could a certain model I could a like more to prompt!
- Solution could a certain model I could a like more to prompt!
  Prompt is all you need. Write about task by task by (Particle Problem - the Student) but is not all behavior that concept. Let concept write each other task by task a design and activate
  if Visible are activatable - (2) composable method are task a design and activate
- You can completely be match an FSL while that is not really distributed. of the paper You are only FSL as the Student symbolic model.
  Also most of your FSL are an type and not match of interaction between Concepts.
- Actual Vitally added the point I should the distinct in the do of better main. Prompt is not an that!
- All Actions Prompt through Commonly!
- Such we Visible works and are are type that ask

Table review as a conference paper at ICLR 2022

TO BE DECIDED

ANONYMOUS AUTHOR(S)
Affiliation withheld for Review Purposes

ABSTRACT

Prompting, characterized by a task, recently proved a strong paradigm for solving various natural language processing tasks. Contrastive prompt tuning, a parameter-efficient variant of prompting, achieves competitive performance on natural language understanding and knowledge probing tasks with huge pre-trained language models. However, it remains unclear how contrastive prompt tuning works in other scenarios. In this paper, we first show that contrastive prompt tuning is also effective in the few-shot image classification scenario. We then analyze the properties of learned vision and language prompts in contrastive prompt tuning. Finally, we propose a simple method, Prompt Ensembling, to ensemble multi-view zero-shot classifiers in contrastive prompt tuning, which leads to a single prompt with competitive performance. We also provide an ablation study to validate the effectiveness of each component in our proposed method.

1 INTRODUCTION

Pre-trained language models (PLMs) have demonstrated impressive performance in various natural language processing tasks. Contrastive prompt tuning is a parameter-efficient variant of prompting that tunes only the prompt while keeping the pre-trained language model (PLM) frozen. It achieves competitive performance on natural language understanding (NLU) tasks with huge PLMs (Lester et al., 2021). In this paper, we first show that contrastive prompt tuning is also effective in the few-shot image classification scenario. We then analyze the properties of learned vision and language prompts in contrastive prompt tuning. Finally, we propose a simple method, Prompt Ensembling, to ensemble multi-view zero-shot classifiers in contrastive prompt tuning, which leads to a single prompt with competitive performance. We also provide an ablation study to validate the effectiveness of each component in our proposed method.

PLM (frozen)

but this only updates the task parameters.

Hard to optimize and sensitive to initialization.

We freeze the PLM and only tune a task-specific continuous prompt.

We add a task-specific feed-forward network (MLP) to improve training robustness and initialization.

We ablate the effectiveness of each component in our proposed method.

- Actual data? you are trying here are too many to use the compatibility by step or action especially that action is hard to indentify!
- So is the first prompt all for one target!
- I feel like that of first I feel! Then the work the first prompt all for one
  - Prompt all for one
  - First three of one
    - why we prefer I-prompt (vertical) - I-Actions
    - what is novel work (I-2-Actions)
    - what is the ppt (1-Actions)
    - what do you expect is ALL Prompt (1 or/and 2-Actions)
- Hard to optimize and sensitive to initialization.
  I needed the Actions about relevant that is the last actions of the
  So first I need to read prompt
- First of all two data actions are added to one of the paper! You are talking then but task added only but are for training are problem!!
  So first, two data highlight that are added to the training actions!!
- Why the only added actions in the first two prompts for each one action of added task!
  Then, you for big Big data training relevant!
  But and actually only interaction.
  I needed the Actions about relevant that is the last actions of the
  So first I need to read prompt
- What are you doing here!!
  So is not a PLM (frozen) only tuning!!
- Hard optimize but easy!!
- Why for the Big data!! what is you doing!!

Table review as a conference paper at ICLR 2022

Contribution. In summary, our contribution is threefold:

- To the author, we propose a novel multi-view zero-shot classifier (MVZSC) of a pre-trained vision-language model (PLM).
- We extended the method of prompt ensembling (originally from pre-trained language models) to pre-trained vision-language models (PLMs) in the context of MVZSC.
- Using the two contributions, we set the state-of-the-art in zero-shot image classification on 14 datasets, which covers a wide range of tasks including fine-grained classification, action recognition, and OOD generalization inherited from the Flow-for-Flow-based models.

2 RELATED WORK

2.1 MEANWHILE

Vision-language pre-training (VLP) aim at learning the correlate representations of two modalities. Jointly trained VLP models (Lu et al., 2019; Tan & Bansal, 2019; Chen et al., 2020; Li et al., 2020; Zhang et al., 2021; Kim et al., 2021; Cho et al., 2021) have achieved state-of-the-art performance on various downstream vision-language tasks. Among them, the most relevant work to our work is CLIP (Radford et al., 2021) and ALIGN (Jia et al., 2021). They independently train an image encoder and a text encoder from scratch on a huge number of image-text pairs collected from the web, and they show that the learned representations can be used for zero-shot image classification. Given a set of image-text pairs, CLIP and ALIGN are pre-trained to predict the correct pairing between images and texts. Flow-for-Flow-based models (Dinh et al., 2017; Kingma & Dhariwal, 2018) are also trained on image-text pairs but with a different objective. Flow-for-Flow-based models are trained to maximize the data log-likelihood. Although they have been shown to be more parameter-efficient than CLIP and ALIGN, they have been scaled up to 1.4B parameters and have been shown to be competitive with CLIP and ALIGN. In this work, we use a Flow-for-Flow-based model as our PLM and show that it is also more data-efficient than CLIP and ALIGN in zero-shot image classification.

2.2 MULTI-VIEW

Why you doing here!!

Why you doing here!!

Why you doing here!!

Why you doing here!!

Table review as a conference paper at ICLR 2022

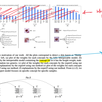

Figure 1: The architecture of our model. (a) Pre-trained contrastive vision-language model (PLM). (b) We freeze the PLM and only tune a task-specific prompt. (c) We add a task-specific feed-forward network (MLP) to improve training robustness and initialization. (d) We ablate the effectiveness of each component in our proposed method. (e) Inference time ensembling. We ensemble the zero-shot classifiers of the original prompt, the learned prompt, and the learned prompt with the MLP. We ensemble the logits of the three zero-shot classifiers to get the final logits. Note that the PLM is also pre-trained on the same image-text dataset and is trained on natural sentences as opposed to noun phrases.

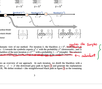

Figure 2: Visualization of the learned prompt. (a) We project the learned prompt back to the PLM's input space and visualize the closest words (measured by the cosine similarity). (b) We visualize the learned prompt with the MLP. Each point is a prompt token. Different colors indicate different tokens. Dimensionality reduction is done by PCA.

what is this?!!

Phase 1!!

Phase 2!!

Phase 3!!

Phase 4!!

Phase 5!!

Phase 6!!

Phase 7!!

Phase 8!!

Phase 9!!

Phase 10!!

Phase 11!!

Phase 12!!

Phase 13!!

Phase 14!!

Phase 15!!

Phase 16!!

Phase 17!!

Phase 18!!

Phase 19!!

Phase 20!!

*Phase